

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

**In re U.S. Patent Application of**)  
SEKI et al.)  
**Application Number: To be Assigned**)  
**Filed: Concurrently Herewith**)  
**For: QUESTION-ANSWERING METHOD AND**)  
**QUESTION-ANSWERING APPARATUS**)  
**ATTORNEY DOCKET NO. ASAM.0103**)

**Honorable Assistant Commissioner  
for Patents  
Washington, D.C. 20231**

**REQUEST FOR PRIORITY  
UNDER 35 U.S.C. § 119  
AND THE INTERNATIONAL CONVENTION**

Sir:

In the matter of the above-captioned application for a United States patent, notice is hereby given that the Applicant claims the priority date of September 12, 2003, the filing date of the corresponding Japanese patent application 2003-320602.

A certified copy of Japanese patent application 2003-320602 is being submitted herewith. Acknowledgment of receipt of the certified copy is respectfully requested in due course.

Respectfully submitted,

Stanley P. Fisher  
Registration Number 24,344

Juan Carlos A. Marquez  
Registration Number 34,072

**REED SMITH LLP**  
3110 Fairview Park Drive  
Suite 1400  
Falls Church, Virginia 22042  
(703) 641-4200  
**January 29, 2004**

日本国特許庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日 Date of Application: 2003年 9月12日

出願番号 Application Number: 特願2003-320602

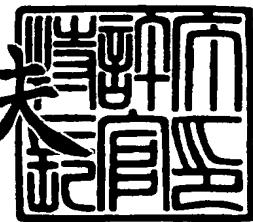
[ST. 10/C]: [JP2003-320602]

出願人 Applicant(s): 株式会社日立製作所

2004年 1月 8日

特許庁長官  
Commissioner,  
Japan Patent Office

今井 康夫



【書類名】 特許願  
【整理番号】 H03009471A  
【あて先】 特許庁長官 殿  
【国際特許分類】 G06F 13/00  
【発明者】  
【住所又は居所】 東京都国分寺市東恋ヶ窪一丁目 280 番地 株式会社日立製作所  
中央研究所内  
【氏名】 関 峰伸  
【発明者】  
【住所又は居所】 東京都国分寺市東恋ヶ窪一丁目 280 番地 株式会社日立製作所  
中央研究所内  
【氏名】 丸川 勝美  
【発明者】  
【住所又は居所】 東京都江東区新砂一丁目 6 番 27 号 株式会社日立製作所公共シ  
ステム事業部内  
【氏名】 ▲高▼津 昌信  
【特許出願人】  
【識別番号】 000005108  
【氏名又は名称】 株式会社日立製作所  
【代理人】  
【識別番号】 100075096  
【弁理士】  
【氏名又は名称】 作田 康夫  
【電話番号】 03-3212-1111  
【手数料の表示】  
【予納台帳番号】 013088  
【納付金額】 21,000円  
【提出物件の目録】  
【物件名】 特許請求の範囲 1  
【物件名】 明細書 1  
【物件名】 図面 1  
【物件名】 要約書 1

**【書類名】特許請求の範囲****【請求項 1】**

質問文書の入力を受け付ける通信装置と、複数の回答例文を記憶する記憶装置と、該複数の回答例文から選択された回答例文を用いて上記質問文書に対する回答作成処理を行うCPUとを有する質問応答装置に、

上記入力された質問文書から複数の重要区間を抽出するステップと、

上記複数の重要区間のそれぞれに対応して求められる上記複数の回答例文の回答例文候補尤度に基づいて上記複数の重要区間を1または複数の重要部に結合するステップと、

上記記憶されている複数の回答例文を用いて上記重要部に対応する回答例文候補を求めるステップとを実行させるためのプログラム。

**【請求項 2】**

請求項1記載のプログラムであって、

上記記憶装置は、上記回答例文のそれぞれについて該回答例文に対応する質問文書に含まれるキーワードの頻度を示す回答例文キーワード頻度情報を記憶し、

上記重要区間を抽出するステップは、

上記入力された質問文書を複数の区間に区切るステップと、

上記回答例文キーワード頻度情報を用いて、上記複数の区間の各区間について、該区間が上記記憶されている複数の回答例文のそれぞれに対応する質問内容である尤度を求めるステップと、

上記各区間について求められた尤度のいずれかと予め定められた閾値との関係に基づいて上記重要区間を抽出するステップとを有することを特徴とするプログラム。

**【請求項 3】**

請求項1記載のプログラムであって、

上記記憶装置は、質問文書の重要部に用いられるキーワードについての重要部キーワード頻度情報または質問文書の非重要部に用いられるキーワードについての非重要部キーワード頻度情報を記憶し、

上記重要区間を抽出するステップは、

上記入力された質問文書を複数の区間に区切るステップと、

上記重要部キーワード頻度情報または上記非重要部キーワード頻度情報を用いて、上記複数の区間の各区間の重要度を計算するステップと、

上記計算された重要度が所定の閾値を超える区間を重要区間として抽出するステップとを有することを特徴とするプログラム。

**【請求項 4】**

請求項1記載のプログラムであって、

上記重要区間を抽出するステップは、第1の閾値を用いた第1の抽出結果と、該第1の閾値よりも低い第2の閾値とを用いた第2の抽出結果とを出力し、

上記重要部に結合するステップと、上記重要部に対応する回答例文候補を求めるステップとは、上記第1及び第2の抽出結果について行われ、

上記第1の抽出結果を用いた場合に求められる回答例文候補と上記第2の抽出結果を用いた場合に求められる回答例文候補との類似度を判断し、

上記類似度に応じて、上記重要部に対応する回答例文候補を求めるステップの処理結果の出力先を、該質問応答装置が通信回線を介して接続される第1及び第2の回答作成端末装置のいずれにするかを決定する振り分けステップとを実行させることを特徴とするプログラム。

**【請求項 5】**

請求項1記載のプログラムであって、

上記重要部に対応する回答例文候補についての振り分け先評価値を計算して所定の閾値と比較するステップと、

上記比較結果に基づいて、上記重要部に対応する回答例文候補を求めるステップの処理結果の出力先を、該質問応答装置が通信回線を介して接続される複数種類の回答作成端末

装置のいずれにするかを決定する振り分けステップとを実行させることを特徴とするプログラム。

**【請求項6】**

請求項1記載のプログラムであって、

上記質問文書と、上記抽出された重要部の箇所を特定する重要部情報と、該重要部に対応する回答例文候補を求めるステップの処理結果である回答例文候補情報を、該質問応答装置が通信回線を介して接続され、表示装置と入力装置とを具備する回答作成端末装置に送信するステップと、

上記回答作成端末装置の表示装置に、上記重要部情報で特定される重要な箇所を第1の強調表示で強調表示した上記質問文書と、上記回答例文候補情報に示される回答例文候補を含む回答文書とを表示するステップと、

上記入力装置を用いて、該表示された質問文書中の一の重要な部を選択する選択入力を受け付けるステップと、

上記選択入力で選択された上記一の重要な部を第2の強調表示で強調表示し、または該回答文書中の上記一の重要な部に対応する回答例文候補を第3の強調表示で強調表示するステップとを実行させることを特徴とするプログラム。

**【請求項7】**

請求項6記載のプログラムであって、

上記送信ステップは、上記各重要な部に対応する回答例文候補を、その回答例文候補尤度が高い順に複数個送信することを特徴とし、

上記一の重要な部を選択する選択入力に応答して、該重要な部に対応する複数の回答例文候補を、上記回答文書とは別に表示するステップとを実行させることを特徴とするプログラム。

**【請求項8】**

請求項7記載のプログラムであって、

上記入力装置において、上記複数の回答例文候補の何れかを選択する選択入力を受け付けるステップと、

上記回答例文候補の選択入力に応答して、上記回答文書中に該選択された回答例文候補を表示するステップとを実行させることを特徴とするプログラム。

**【請求項9】**

請求項6記載のプログラムであって、

上記表示された重要な部中の所定の文字列を強調表示させることを特徴とするプログラム。

**【請求項10】**

請求項6記載のプログラムであって、

上記各重要な部を異なる強調表示で表示させることを特徴とするプログラム。

**【請求項11】**

請求項6記載のプログラムであって、

上記各重要な部と合わせて該重要な部のID情報を表示させることを特徴とするプログラム。

**【請求項12】**

質問応答装置と、ネットワークを介して該質問応答装置と接続される回答作成端末装置とを有する質問応答システムであって、

上記質問応答装置は、質問文書の入力を受け付ける通信装置と、複数の回答例文を記憶する記憶装置と、該複数の回答例文から選択された回答例文を用いて上記質問文書に対する回答作成処理を行う処理部とを有し、

上記回答作成端末装置は、該回答作成処理結果を受信する通信装置と、該回答作成処理結果に含まれる情報を表示する表示装置と、該表示情報に対する入力を受け付ける入力装置とを有し、

上記質問応答装置の処理部は、

上記入力された質問文書から複数の重要な区間を抽出し、上記複数の重要な区間のそれぞれ

に対応して求められる上記複数の回答例文の回答例文候補尤度に基づいて上記複数の重要区間を1または複数の重要な部に結合し、上記記憶されている複数の回答例文を用いて上記重要な部に対応する回答例文候補を求めて回答作成処理結果を出力することを特徴とする質問応答システム。

#### 【請求項13】

請求項12記載の質問応答システムであって、

上記質問応答装置の記憶装置は、上記回答例文のそれぞれについて該回答例文に対応する質問文書に含まれるキーワードの頻度を示す回答例文キーワード頻度情報を記憶し、上記重要区間の抽出は、

上記入力された質問文書を複数の区間に区切り、

上記回答例文キーワード頻度情報を用いて、上記複数の区間の各区間について、該区間が上記記憶されている複数の回答例文のそれぞれに対応する質問内容である尤度を求め、

上記各区間について求められた尤度のいずれかと予め定められた閾値との関係に基づいて上記重要区間を抽出することにより実行されることを特徴とする質問応答システム。

#### 【請求項14】

請求項12記載の質問応答システムであって、

上記記憶装置は、質問文書の重要な部に用いられるキーワードについての重要な部キーワード頻度情報をまた質問文書の非重要な部に用いられるキーワードについての非重要な部キーワード頻度情報を記憶し、

上記重要区間の抽出は、

上記入力された質問文書を複数の区間に区切り、

上記重要な部キーワード頻度情報または上記非重要な部キーワード頻度情報を用いて、上記複数の区間の各区間の重要度を計算し、

上記計算された重要度が所定の閾値を超える区間を重要区間として抽出することで実行されることを特徴とする質問応答システム。

#### 【請求項15】

請求項12記載の質問応答システムであって、

上記重要区間を抽出する際に、第1の閾値を用いた第1の抽出結果と、該第1の閾値よりも低い第2の閾値とを用いた第2の抽出結果とを出力し、

上記重要な部に結合する処理と、上記重要な部に対応する回答例文候補を求める処理とは、上記第1及び第2の抽出結果について行い、

上記第1の抽出結果を用いた場合に求められる回答例文候補と上記第2の抽出結果を用いた場合に求められる回答例文候補との類似度を判断し、

上記類似度に応じて、上記重要な部に対応する回答例文候補を求めるステップの処理結果の出力先を、該質問応答装置が通信回線を介して接続される第1及び第2の回答作成端末装置のいずれにするかを決定して振り分けることを特徴とする質問応答システム。

#### 【請求項16】

請求項12記載の質問応答システムであって、複数の種類の上記回答作成端末装置を有し、

上記質問応答装置の処理部は、

上記重要な部に対応する回答例文候補についての振り分け評価値を計算して所定の閾値と比較し、

上記比較結果に基づいて、上記重要な部に対応する回答例文候補を求めるステップの処理結果の出力先を、上記複数の種類の回答作成端末装置のいずれにするかを決定する振り分けを実行することを特徴とする質問応答システム。

#### 【請求項17】

請求項12記載の質問応答システムであって、

上記質問応答装置は、上記質問文書と、上記抽出された重要な部の箇所を特定する重要な部情報を、該重要な部に対応する回答例文候補を求めるステップの処理結果である回答例文候補情報を上記回答作成端末装置に送信し、

上記回答作成端末装置はその表示装置に、上記重要部情報で特定される重要な箇所を第1の強調表示で強調表示した上記質問文書と、上記回答例文候補情報に示される回答例文候補を含む回答文書とを表示することを特徴とする質問応答システム。

【請求項18】

請求項17記載の質問応答システムであって、

上記回答作成端末装置は、上記入力装置を介して該表示された質問文書中の一の重要な部を選択する選択入力を受け付け、

上記表示装置は、上記選択入力で選択された上記一の重要な部を第2の強調表示で強調表示し、または該回答文書中の上記一の重要な部に対応する回答例文候補を第3の強調表示で強調表示することを特徴とする質問応答システム。

【請求項19】

請求項18記載の質問応答システムであって、

上記質問応答装置は、上記各重要な部に対応する回答例文候補を、その回答例文候補尤度が高い順に複数個送信し、

上記表示装置は、

上記一の重要な部を選択する選択入力に応答して、該重要な部に対応する複数の回答例文候補を、上記回答文書とは別に表示し、

上記入力装置を介して入力される、上記複数の回答例文候補の何れかを選択する選択入力に応答して、上記回答文書中に表示されていた回答例文候補を、該選択された回答例文候補で置き換えて表示することを特徴とする質問応答システム。

【請求項20】

請求項17記載の質問応答システムであって、上記各重要な部を異なる強調表示で表示することを特徴とする質問応答システム。

【書類名】明細書

【発明の名称】質問応答方法及び質問応答装置

【技術分野】

【0001】

本発明は、テキスト化された質問文書に対して、回答する質問回答処理を支援する自動回答方法及び自動回答装置に関する。

【背景技術】

【0002】

近年の電子化の普及に伴い、企業などへの質問に、ホームページのフォーム入力や、電子メールが利用されるケースが多くなっている。しかし、これら質問1つ1つに対して、企業側が人手で対応しようとすると、多くのオペレータを必要としコストが大きくなってしまう。また経験の浅いオペレータでは回答できない、或いは回答するのに非常に多くの時間を費やしてしまう場合がある。これに対して、質問文書を入力し、その内容を解析し、予め用意しておいた質問内容毎の回答例文や質問回答事例集の中から、回答例文候補を選択し提示する質問応答システムが導入され始めている。

【0003】

ただしこれら質問応答システムの多くは、1つの文書内には1つの相談内容があることを想定しており、1つの文書内に複数の質問内容が記載された場合には個々の質問内容をそれぞれ解析することができず、回答精度が低い。

【0004】

これに対して、特開2002-132661号公報(特許文献1)に開示された技術がある。これは質問内容が複数含まれる場合、その内容毎に質問内容を分割する手段をもち、分割された内容それぞれについて解析し、回答例文候補を選択する。またそれぞれの回答例文候補の正しさの度合を示す回答精度を算出し、その回答精度が所定値以上であれば、回答例文候補から回答を作成し、回答精度が所定値以下であれば回答を新規に作成することを促す指示をする。

【0005】

【特許文献1】特開2002-132661号公報

【0006】

【非特許文献1】北研二, 津田和彦, 獅々堀正幹著、「情報検索アルゴリズム」、共立出版、2002年1月

【発明の開示】

【発明が解決しようとする課題】

【0007】

しかし、特許文献1に開示された従来技術では、文書を内容毎に分割する手段は「(数字)」、「(英字)」、「・」や、字下げ、「ところで」等の接続詞を用いて分割処理を行うと記されているが、質問内容内に、「(数字)」や字下げなどを区切りとして文書を分割してしまうと、質問内容を複数に分割してしまうという問題があった。また逆に質問内容の範囲を大きく取り過ぎてしまい、複数の質問内容に対して、それぞれ回答例文候補を選択することができないという問題があった。

【0008】

また従来技術では質問文書の情報のみを用いて文書を分割するため、その分割された範囲が回答例文候補では対応できない場合がある。すなわち質問内容だけをみて分割するのではなく、用意された回答例文候補に対応するように文書を分割する必要がある。

【0009】

また従来技術で分割された文書の中には質問内容とは関係のない文も含まれるため、それらが回答例文候補生成処理に悪影響を与え回答例文候補生成精度が低くなる。そして、さらに回答作成するオペレータが文書を見たときに何処の部分を読んだら良いかを探すのに時間がかかるという問題があった。

【0010】

また従来技術では、回答例文候補の正しさの度合を示す回答精度を算出し、その回答精度が所定値以上であれば、回答例文候補から回答を作成し、自動応答(自動返信)する。そして、回答精度が所定値以下であれば回答を新規に作成することを促す指示を出していた。しかし、回答の種類が非常に大きい場合や、似たような質問だが異なる回答を必要とする場合等は、回答精度は低くなり、所定の閾値を超えるサンプルが少なくなる。ゆえに自動応答できるサンプル数が少なく、新たに回答を作成する場合が多くなり、オペレータ業務の効率化が低くなってしまう、或いは誤った回答例文候補を用いて自動応答してしまうという問題があった。

#### 【0011】

この発明は上記のような問題点を解決するためになされたもので、1つの文書内に複数の質問内容がある場合にも、それぞれの質問に対して精度良く回答例文候補を選択する質問応答方法および質問応答装置を提供することを目的とする。

#### 【0012】

また、オペレータに対して質問文書内の重要な部分を提示することで、必要なところのみを読ませることが可能な質問応答装置を提供することを目的とする。

#### 【0013】

また、回答の種類が非常に大きい場合や、似たような質問だが異なる回答を必要とする質問があった場合でも、誤った回答例文を用いて自動返信してしまうことなく、オペレータ業務の効率化が高くなる自動応答方法および自動応答装置を提供することを目的とする。

#### 【課題を解決するための手段】

#### 【0014】

上記問題を解決するために、予め複数の回答例文を用意しておき、質問文書が入力されると、その文書から質問内容毎に重要部を抽出し、重要部毎に予め用意された回答例文から回答例文候補を選択する。

#### 【0015】

これにより、質問内容とは関係のない部分を除き、重要な部分のみを用いて回答例文候補生成を行うため、複数の質問内容がある場合でも回答例文候補の選択精度が高くなる。また、前記重要部の抽出は、文書を所定の区間に区切り、区切られた区間毎にそれぞれ重要であるか否かを判定することで、重要区間を抽出し、抽出された区間毎に各回答例文に対する質問内容であるかの度合を表す回答例文候補尤度を算出し、回答例文候補尤度を用いて意味的に近い重要区間を結合することで重要部を抽出することを特徴とする。

#### 【0016】

これにより、文書から複数の重要部を抽出する際、予め用意された回答例文候補に対応するように重要部をまとめ抽出するため、個々の質問内容に対する回答例文候補の選択精度が高くなる。

#### 【0017】

重要区間の抽出は、回答例文に対する質問内容であるかを表す度合が高く、質問内容の重要な部分であるかを表す度合が高い区間、或いは、上記2つの度合のうちいずれかが高い区間を抽出する。事前に収集された質問文書集合を対応する回答例文毎に分けてデータベースを作成し、回答例文毎に所定のキーワードの出現頻度情報（回答例文キーワード頻度情報）を作成し、また重要部と非重要部に分けてデータベースを作成し、重要部データベースと非重要部データベース、それれにおいて所定のキーワードの頻度情報（重要部キーワード頻度情報、非重要部キーワード頻度情報）を作成し、それら頻度情報を用いて重要であるか否かを判定することを特徴とする。

#### 【0018】

また上記で説明した回答例文候補の選択方法に加え、重要部毎に生成された回答例文候補の確からしさの度合を示す回答例文候補信頼度を算出し、重要部毎に回答例文候補信頼度を用いて回答例文候補が確定であるか否かを判定し、重要部毎に選択された回答新規作成が必要か否かの度合を示す回答新規作成度を算出し、重要部毎に回答新規作成度を用い

て回答新規作成が必要であるか否かを判定し、すべての重要部において回答例文候補を確かに正しいと判定した場合と、1つ以上の重要部において回答新規作成が必要であると判定した場合と、それ以外の場合で、異なるオペレータ端末に質問文書を振り分けることを特徴とする。

#### 【0019】

これにより、すべての重要部において回答例文候補を確定すると判定した文書は、直接決裁者端末に送ることが可能となり、この文書を受取った決裁者端末では、重要部毎に、それらの回答例文候補が正しいかどうかを確認するのみの作業(回答確認作業)が行われ、すべての質問内容に対する回答が正しいことが確認されると、質問者に作成した回答が返信される。そして、1つ以上の重要部において回答新規作成が必要であると判定した文書とその回答例文候補を受取った高知識オペレータ端末では、回答新規作成と判定された重要部に対して、回答例文新規作成作業を行い、その他の重要部に対しては、重要部毎に1つ以上ある回答例文候補の中から回答が選択されて、回答を作成し、決済者端末に送る。そして、それ以外の文書とその回答例文候補を受取った一般オペレータ端末では、重要部毎に1つ以上ある回答例文候補の中から回答を選択して、回答が作成され、決済者端末に送る。

#### 【0020】

これにより、従来すべての文書に対して重要部毎に、回答選択作業、或いは回答新規作成を行わなければならなかったが、直接決済者端末に送られる文書は、回答候補の確認作業を行うだけで済むようになり、業務が効率化される。またオペレータの数が削減され、コスト削減されるそして、一般オペレータにより、回答を選択するだけで済む文書の回答作成作業が行われ、高知識オペレータにより、回答新規作成を必要とする質問文書の回答作成作業が行われるため、業務が効率化される。そして、一般的に賃金の高い高知識オペレータの数を削減でき、コスト削減できる。

#### 【発明を実施するための最良の形態】

#### 【0021】

はじめに実施の形態1において、本願発明の基本構成の実施を行った場合を説明し、実施の形態2において、基本構成にオペレータへの自動振り分けを加えた構成について説明する。

#### 【実施例1】

#### 【0022】

以下、図を用いて本願発明を説明する。図1は本願発明の基本構成図である。図1において、101文書入力ではテキスト化された文書が入力される。このテキスト化された文書はEメールであったり、紙文書をOCRで読み取ったものであったり、様々な場合がありえる。図3はテキスト化された文書の例であり、句点を除き内容は全て“—”で表してある。

#### 【0023】

そして、重要部の抽出を行う。ここでは、入力された文書から質問内容毎に重要部を抽出する。

#### 【0024】

まず入力された文書に対して、処理区間の決定102を行う。ここでは、文書中の改行、句点を境として文書を区切る。図4にその例を示す。

#### 【0025】

次に重要区間の抽出105を行う。ここでは、処理区間の決定102で区切られた区内の文章が、図11に示すような回答例文毎の質問文書集合とそれに対応した回答例文用キーワード頻度情報（例：1104、1105、1106）と、図10に示す重要部と非重要部に分けられた文書集合とそれに対応する重要部キーワード頻度情報（901）と非重要部キーワード頻度情報（902）を用いる。

#### 【0026】

まず回答例文用キーワード頻度情報の作成方法について図11を用いて説明する。回答

例文が3つの場合の例であるが、一般には数百の回答例文を用いる。以降この3つの回答例文を用いる場合として説明していく。まず事前に複数の質問サンプル文書をそれぞれ対応する回答例文毎に分類する。一つのサンプル文書が複数の回答例文を必要とする場合には、それらの文書を分離し、それぞれが一つの回答例文で対応できるように加工する。そして分離されたそれぞれの文書を対応する回答例文毎に分類する。そして分類された文書集合毎、すなわち回答例文毎に、予め設定されたL個のキーワード( $W_1, W_2, \dots, W_L$ )の頻度情報を作成する。回答例文 A 001 に対しては ( $A001W_{c1}, A001W_{c2}, \dots, A001W_{cL}$ ) , 回答例文 B 002 に対しては ( $B002W_{c1}, B002W_{c2}, \dots, B002W_{cL}$ ) , 回答例文 C 003 に対しては ( $C003W_{c1}, C003W_{c2}, \dots, C003W_{cL}$ ) がキーワード頻度情報となる。

### 【0027】

次に重要部キーワード頻度情報と非重要部キーワード頻度情報の作成方法を図12を用いて説明する。まず複数のサンプル文書を重要部と非重要部に分ける(重要部集合1212、非重要部集合1213)。そして重要部集合、非重要部集合、それに対して予め設定されたL個のキーワード( $W_1, W_2, \dots, W_L$ )の頻度情報を作成する。重要部に対しては ( $IW_{c1}, IW_{c2}, \dots, IW_{cL}$ ) 、非重要部に対しては ( $UW_{c1}, UW_{c2}, \dots, UW_{cL}$ ) がキーワード頻度情報となる。キーワード頻度情報は、予め設定されたキーワードが与えられた文書集合中に出現した回数であり、例えば ( $A001W_{c1}, A001W_{c2}, \dots, A001W_{cL}$ ) は、A001の回答例文に対応する質問文書の集合に各キーワードが何回出現したかを表す。ただし、これらは単純にキーワードの出現回数とは限らず、対応する質問文書集合中のキーワードの出現傾向を表すものであれば良い。またキーワード( $W_1, W_2, \dots, W_L$ )は、単語であるとは限らず、单文字や、一文内に共起する単語のペア等、文書を特徴付ける要素であれば良い。

### 【0028】

以上が各キーワード頻度情報についての説明である。そして、これらキーワード頻度情報を用いて重要区間の抽出をする方法を説明する。

### 【0029】

重要区間の抽出は、処理区間決定102で区切られたそれぞれの区間毎に処理を行う。まず所定のキーワード( $W_1, W_2, \dots, W_L$ )が区間に何回出現したかを数える( $PW_1, PW_2, \dots, PW_L$ )。そして、区間毎の回答候補重要度の算出401にて、各回答例文に対する区間回答候補重要度 ( $A001LhdIU, B002LhdIU, C003LhdIU$ ) を算出する。区間回答候補重要度は回答例文毎にそれぞれ以下の計算式で算出する。ただし、方法は以下の計算式に限らず、各区間の文章が各回答例文に対する質問内容であるかを表す度合を示す計算式、或いは質問内容として重要内容であるかを表す度合を示す計算式、或いは各回答例文に対する質問内容であり重要内容であるかを表す度合を示す計算式であれば良い。例えば、重要部キーワード頻度情報、非重要部キーワード頻度情報、回答例文キーワード頻度情報のいずれかを除いた計算式でも良い。また共立出版、北研二、津田和彦、獅々堀正幹著、情報検索アルゴリズム（非特許文献1）に記載されているように、例えば、IDF（Inverse Document Frequency）算出による方法などもある。以下の式は重要区間である可能性が高いほど大きい値をとるが、逆に重要区間である可能性が高いほど小さい値をとる計算式の場合もある。

### 【0030】

#### 【数1】

$$\begin{aligned} A001LhdIU &= \sum_{i=1}^L PW_i \times A001W_{ci} \times IW_{ci} / UW_{ci} \\ B002LhdIU &= \sum_{i=1}^L PW_i \times B002W_{ci} \times IW_{ci} / UW_{ci} \\ C003LhdIU &= \sum_{i=1}^L PW_i \times C003W_{ci} \times IW_{ci} / UW_{ci} \end{aligned} \quad \text{数1}$$

### 【0031】

を計算する。そして、回答例文候補重要度の高い順に回答例文をソートする。

#### 【0032】

そして、回答例文候補重要度による重要区間の抽出402において、区間毎に区間回答候補重要度を予めカテゴリ毎に設定された閾値と比較し、一つ以上の回答例文が閾値以上であれば、その区間を重要区間として抽出する。上記計算式が重要区間が高いほど低い値をとる場合には、予めカテゴリ毎に設定された閾値と比較し、一つ以上の回答例文が閾値以下であれば、その区間を重要区間として抽出する。

#### 【0033】

図5にその表示例を示す。網掛けで表示されている部分が、重要区間である。この重要区間の抽出方法は一例である。処理区間決定102により決まる区間を固定文字数にすることや、予め設定されたキーワードを境に区切ることもできる。

#### 【0034】

このように、回答例文キーワード頻度情報を用いることにより、各回答例文に対応する質問文書の特徴的な部分を抽出できる。そして、重要部キーワード頻度情報を用いることにより、質問文書として重要な区間を抽出できる。そして、非重要部キーワード頻度情報を用いることにより、重要でない区間の抽出を抑制、すなわち重要な区間の抽出精度を良くすることができる。以上のように、回答例文キーワード頻度情報や重要部キーワード頻度情報や非重要部キーワード頻度情報を用いることにより、固定したキーワードや表現を含む文に限らず、重要箇所を抽出できる。またこれらを組み合わせて用いることにより、より精度の高い重要区間の抽出ができる。

#### 【0035】

次に、重要区間の結合104を行う。ここでは、重要区間として抽出した複数の区間の中から、同じ回答例文を用いる部分、すなわち同じ質問内容に関する部分を結合する。

#### 【0036】

まず、重要区間毎に各回答例文に対する区間回答候補尤度(A001Lhd, B002Lhd, C003Lhd)を算出する。区間回答候補尤度は回答例文毎にそれぞれ以下の計算式で算出する。ただし、方法は以下の計算式に限らず、各区間の文章が各回答例文に対する質問内容であるかを表す度合を示す計算式であれば良く、例えば重要区間の抽出に用いた値をそのまま用いても良い。

#### 【0037】

##### 【数2】

$$\begin{aligned} A001LhdIU &= \sum_{i=1}^L PW_i \times A001Wc_i \\ B002LhdIU &= \sum_{i=1}^L PW_i \times B002Wc_i \\ C003LhdIU &= \sum_{i=1}^L PW_i \times C003Wc_i \end{aligned} \quad \text{数2}$$

#### 【0038】

そして、回答例文候補尤度の高い順に回答例文をソートする。そして、回答例文候補尤度の高さが第1位の回答例文候補が同じである重要区間同士を結合して重要部を抽出する。さらに、結合した区間において、再度回答例文選択尤度を算出し、回答例文候補尤度の高い順に処理を行い、その結果、第1位の回答例文候補が同じ重要区間が存在すれば、それらを結合する。これらを第1位の回答例文候補が同じ重要区間がなくなるまで繰り返すこともできる。また2位、3位の回答例文候補の結果を考慮して結合処理を行うこともできる。また重要区間の始まりが“したがって”や“それ”など前の文を受けるようなキーワードで始まる場合に、前に隣接する重要区間があれば、その重要区間と結合を行うこともできる。

**【0039】**

以上が重要部の抽出処理例である。共通の回答例文候補をもつ重要区間どうしを結合して重要部をまとめることにより、質問内容を複数に分割することや、逆に複数の質問内容にまたがる範囲を抽出してしまうことを防ぐことができる。あるいは、質問文書の内容が比較的単純な場合には、改行ごとに文書を区切り、予め設定されたキーワードを含む部分を抽出するなどの簡単な処理とすることもできる。

**【0040】**

次に、重要部毎の回答例文候補生成105を行う。ここでの処理は、求めた重要部がそれぞれ各回答例文に対する質問内容であるかを表す度合を計算し、その値が高い順に上位X件の回答例文候補を選択する。Xの値はユーザが自由に設定可能である。この際、選択された回答例文に穴埋め箇所が存在する場合は、質問文書に付加された質問者ID等をキーとして、穴埋め箇所に対応する情報を穴埋め用データベースから検索し、穴埋めされた回答例文を作成し、これを回答例文候補とする。

**【0041】**

まず、重要部毎に回答候補尤度を算出し、回答候補尤度の値が高い順に上位X件を選択する。回答候補尤度の算出は重要区間の結合104に用いた方法と同じである。そして重要部毎に行う。算出する文章の単位が異なるだけで処理方法は同じことを行う。

**【0042】**

これにより、質問内容とは関係のない部分を除き、重要な部分のみを用いて回答例文候補生成を行うため、複数の質問内容がある場合でも回答例文候補精度が高くなる。また従来技術では質問文書の情報のみを用いて文書を分割するため、その分割された範囲が回答例文候補では対応できない場合があるが、文書から複数の重要部を抽出する際、予め用意された回答例文候補に対応するように重要部をまとめ抽出するため、用意された回答例文候補に対応するように重要部を抽出することが可能であり、個々の質問内容に対する回答例文候補生成精度が高くなる。

**【0043】**

これら一連の処理(101～105、401～402)は、図13に示すような文書自動処理1301で実行される。文書自動処理装置は例えばネットワークに接続されたコンピュータ、サーバ装置等により実現される。そして、これら一連の処理を行うプログラムは外部記憶装置1304やメモリ1305等の記憶装置に格納され、CPU(Central Processing Unit)によって実行される。文書入力101は、CPU(Central Processing Unit)により、質問文書が通信装置1302、キーボードやOCR等の入力装置1303を通して外部記憶装置やメモリに格納されることで実現される。また、回答例文キーワード頻度情報、重要部キーワード頻度情報、非重要部キーワード頻度情報、回答例文、穴埋め用データベース、閾値情報は、プログラム実行中の中間データは外部記憶装置やメモリ等の記憶装置に格納しておき、図1に示す一連の処理に従ってCPUによりロード/ストアされる。そして、一連の処理の結果は、外部記憶装置やメモリ等の記憶装置に格納され、液晶ディスプレイやCRT等の表示装置1304に表示される、または、通信装置1302を通して外部へ出力され、通信回路を介して接続される他の装置の表示装置に表示される。

**【0044】**

上記のように一連の処理が自動処理装置によって実行された結果、質問内容ごとに複数の重要部とその質問内容に対する回答例文候補が選択される。

**【0045】**

そして、得られた重要部とその回答候補の結果が表示装置を用いて提示される。質問応答装置から回答作成オペレータ等のための端末に応答作成結果を転送して表示させる場合には、質問文書、重要部情報(重要部の箇所、対応回答例文候補などを示す情報)などが、質問応答装置から端末に送信される。上記の情報のほか、各重要部に対応する複数の回答例文候補の順位情報や尤度情報、重要部中のキーワード箇所を示す情報、等をあわせて送信してもよい。これらの情報は、質問応答装置または端末の表示装置で回答作成オペレータに提示され、回答作成オペレータはこれを用いて回答を作成し、決済者に作成結果を

送信、或いは直接質問者に返信をする。これを行う際の回答作成オペレータの回答作成画面、すなわち重要部とその回答候補の表示画面について図15を用いて説明する。回答作成画面1507は、質問/相談文書表示部1501と回答作成部1502と回答例文候補表示部1503と宛先表示/編集部1504と送信ボタン1506と重要部のみを表示させるためのボタン、重要部のみ表示ボタン1505を有する。

#### 【0046】

質問/相談文書表示部1501は質問者からのテキスト化された文書が表示される。この際、抽出された重要部1508、1509はそれ以外とはことなり背景に色づけされるなどの強調表示がなされる。また、重要部毎にID1510、1511がふられ、各重要部の隣に表示される。また、各重要部を、異なる背景色や異なる枠で囲むなどの、異なる強調表示をする。これにより、相談文書中に重要部が複数あっても、それぞれの存在を明らかに見分けられる。また文書全体の閲覧が容易である。また重要部として抽出した箇所の近傍の文章も見やすく、重要部の抽出箇所にずれがある場合にも回答作成が容易である。また重要部が抽出されなかった場合にも、回答作成作業を行うことができる。また、複数の重要部を一度に閲覧できることにより、回答作成者が複数の重要部の関係を把握しやすくなるという利点もある。例えば、一つの重要部として抽出されるべきであるにもかかわらず複数の重要部として抽出されてしまった場合にその間違いを容易に把握して訂正できる、また、前後の重要部を一度に閲覧することにより、回答作成者が文脈を考慮してより適切な回答例文を選択することが容易になる、等の効果がある。

#### 【0047】

オペレータが使用する端末には、マウスやキーボード等の入力装置が備えられ、オペレータは、マウスを用いて表示画面中の各表示オブジェクトの選択、表示内容・表示方法の変更などを行うことができ、また、キーボードを用いて回答文の作成・追加などの編集を行うことができる。

#### 【0048】

また重要部のみ表示ボタン1505を押す（マウス等を用いて選択する）と質問/相談文書表示部1501中には重要部のみが表示され（図12）、もう一度押すとともに戻る。

#### 【0049】

これにより、回答作成するオペレータは、複数存在する質問内容がどこにあるかを瞬時に判別し、重要な部分を効率的に参照することができるため、業務を効率化できる。

#### 【0050】

また予め回答例文毎に複数の文字列をキーワードとして設定しておき、質問/相談文書表示部1501中の各重要部では、それぞれの1位の回答例文候補で設定されているキーワードが他の文字列と異なる文字色で強調表示される。

#### 【0051】

また回答例文候補表示部1503では、回答例文候補の上位5件の順位と回答IDと尤度と説明が表示される。ただし順位表示部分は単なる表示だけでなく、それぞれがボタンとなっている。はじめは、質問/相談文書表示部1501に表示されている質問文書の重要部のうち、最も上に存在する重要部に対する回答例文候補が表示される。そして質問/相談文書表示部1501中の各重要部はその表示部分をクリックすることで選択され、選択された重要部に対する回答例文候補が回答例文候補表示部に表示される。この際、選択された重要部は他の重要部とは異なる背景色で強調表示される。そして、順位ボタンの一つを押すと押された順位に対応する回答例文が回答作成部1502に表示される。また同時に、それまで異なる色で強調表示されていたキーワードの色は他の文字列の色と等しくなり、押された順位に対応する回答例文に対して設定されていたキーワードが他の文字列の色とは異なる色で強調表示される。

#### 【0052】

回答例文候補表示部で何れかの回答例文候補を選択すると、その回答例文候補が回答作成部に表示される。この際に、新しく選択された回答例文候補は、それまで回答作成部に

表示されていた回答例文候補とあわせて表示されてもいいし、それまで表示されていたものと置き換えられてもよい。また、回答作成部に表示中の回答例文候補に対応する項目を回答例文候補表示部で選択すると、その回答例文候補を回答作成部から削除するようにしてもよい。

#### 【0053】

これらの機能を利用して回答例文作成オペレータは回答作成部にて回答を作成する。回答作成部1502では、回答例文候補表示部の回答IDをマウス操作等により選択し挿入された回答例文を用いて回答が作成される。この際、挿入された回答例文をそのまま回答とするだけでなく、キーボード、マウス等の入力装置を用いて、回答例文を編集する、回答例文で対応できない質問内容に対して新たに回答を作成する、作成した部分を訂正/削除する、ことができる。これにより回答作成部にて複数の質問に対する回答を編集できる。

そして、複数の質問内容に対する回答が完成したら宛先表示/編集部1504で宛先を設定し、送信ボタン1506を押し、作成した回答を設定した宛先に送信する。

#### 【0054】

これにより、回答作成するオペレータは、複数抽出された重要部の回答候補と対応する特徴的な文字列を効率的に閲覧することができ、すばやく的確な回答例文を選択して組み合わせ、回答を作成することができるため、業務が効率化できる。

#### 【実施例2】

#### 【0055】

実施例1の構成にオペレータへの自動振り分け機能を追加した場合を説明する。図14は自動振り分け機能を追加した場合のシステム構成を表す図である。文書入力手段1401、自動文書処理装置1402、文書出力手段1403、一般オペレータ端末1405、1404、1405、専門知識オペレータ端末1407、1408、決裁者端末1409、1410がネットワークを介して接続されたシステムであり、それぞれの装置、端末は1つとは限らず、複数存在してもよい。

#### 【0056】

文書入力手段は、質問文書を入力する手段を有する装置であり、メールを受信するメールサーバや紙文書をテキストデータ化するOCR等である。また文書入力手段と文書自動処理装置は1つの装置として構成される場合もある。

#### 【0057】

文書自動処理装置は、図13に示すようなネットワークに接続されたコンピュータ、サーバ装置であり、文書入力装置から質問文書を受取り、図20に示す複数質問毎に回答候補を選択する等の文書処理（自動振り分け処理）を行い、自動振り分け処理結果と質問文書は、自動振り分け処理で得られた結果をもとに、一般オペレータ端末、専門知識オペレータ端末、決裁者端末の中で、いずれかの端末に送られる。

#### 【0058】

一般オペレータ端末では、文書自動処理装置、決裁者端末から送られてきた質問文書と自動振り分け処理結果を入力として、オペレータの動作により回答作成作業が行われ、その結果が、決裁者端末、専門知識オペレータ、他の一般オペレータの中で、オペレータによって指定した端末に送られる。ここには、主に回答例文のみで対応できる質問文書が送られ、オペレータによる回答例文選択作業により、回答が作成される。そして回答例文で対応できない質問文書が送られてきた場合には、質問文書と自動振り分け処理結果は高知識オペレータ端末に送られる。

#### 【0059】

専門知識オペレータ端末では、文書自動処理装置、専門知識オペレータ、決裁者端末、他の一般オペレータ端末から送られてきた文書と自動振り分け処理結果を入力として、オペレータの動作により回答作成作業が行われ、その結果が、決裁者端末、一般オペレータ、他の専門知識オペレータの中で、オペレータによって指定した端末に送られる。ここには、主に回答例文では対応できない質問内容を含む文書が送られ、オペレータにより新規

に回答が作成されることが多くなる。そのためこの作業を行うオペレータは新たな質問内容に答えるだけの高い知識を持っていることが望まれる。

#### 【0060】

決済者端末では、文書自動処理装置、専門知識オペレータ、決済者端末、他の一般オペレータ端末から送られてきた質問文書と回答作成結果を入力として、決裁者により回答作成結果の確認、訂正等が行われる。そして、回答作成が正しいことを確認されれば、文書出力手段に回答が送られる。しかし回答作成が正しくない場合は、再度回答作成が行われるよう指示を加えて、質問文書と自動振り分け処理結果が、一般オペレータ端末、或いは専門知識オペレータ端末に送られる。

#### 【0061】

文書出力手段は、回答を出力する手段を有する装置であり、メールを送信するメールサーバや回答を印刷するプリンタ等であり、この装置をとおして回答が質問者に送られる。また文書出力手段と文書自動処理装置は1つの装置として構成される場合もある。

#### 【0062】

次に図15を用いて、文書自動処理装置の処理を説明する。まず、前述した5つのステップ、文書入力、処理区間の決定、重要区間の抽出、重要区間の結合、重要部毎の回答候補生成を実行する。

#### 【0063】

次に回答例文候補の検定1701を行う。ここでは、重要部毎に選択した第1位の回答例文候補が確かに正しいかどうかを検定する。まず回答例文候補の確からしさの度合を示す回答例文候補信頼度を算出する。これには第1位の回答例文候補の回答例文候補尤度（重要部毎の回答例文候補生成において算出した回答例文候補尤度）をそのまま用いるが、これに限らず回答例文候補の確からしさの度合を示す値であればよく、非特許文献1に記載のようにIDFによる方法で新たに回答例文候補信頼度を算出してもよい。そして、重要部毎に回答例文候補信頼度が予め定められた閾値よりも大きいかどうか比較する。回答例文候補信頼度が閾値よりも大きければ、1位の回答例文候補を確かに正しいと判定し、回答例文候補信頼度が閾値以下であれば、1位の回答例文候補は不確かであると判定する。閾値の値は、質問文書サンプルを用いて回答例文候補信頼度の算出を行った実験結果をもとに、重要部の1位の回答例文候補が正しい場合に閾値より大きい回答例文候補信頼度を持つことが多くなるように設定される。

#### 【0064】

回答例文候補が確かに正しいかどうかを判定する方法は、上記に限らず回答例文毎に複数のキーワードを設定しておき、1位の回答例文候補に対応するキーワードの一部が重要部に含まれている場合には、1知の回答例文候補を確かに正しいと判定し、それ以外の場合は、1位の回答例文候補は不確かであると判定してもよい。

#### 【0065】

次に回答新規作成判定1801を行う。ここでは、回答例文候補では対応できなく新規に回答を作成する必要があるか否かを判別する。まず重要部毎に新規に回答を作成する必要があるか否かの度合を示す回答新規作成度を算出する。これには第1位の回答候補の回答尤度をそのまま用いるが、これに限らず重要部毎に新規に回答を作成する必要があるか否かの度合を示す値であればよく、非特許文献1に記載のようにIDFによる方法で新たに回答例文候補信頼度を算出してもよい。

#### 【0066】

そして、重要部毎に回答新規作成度が予め定められた閾値よりも大きいかどうか比較する。回答新規作成度が閾値よりも小さければ、回答を新規作成する必要があると判定し、回答新規作成度が閾値以上であれば、回答を新規作成する必要があるかどうかは不確かであると判定する。

#### 【0067】

閾値の値は、質問文書サンプルを用いて回答新規作成度の算出を行った実験結果をもとに、重要部に対して回答を新規に作成する必要がある場合に閾値より小さい回答新規作成度を

持つことが多くなるように設定される。

#### 【0068】

次に振り分けC102を行う。ここでは、回答例文候補検定1701で全ての重要部の1位の回答例文候補が確かに正しいと判定した文書とこれまでの一連の処理結果を決裁者端末に送り、回答新規作成1801で一つでも回答新規作成する必要があると判定された重要部が存在する文書とこれまでの一連の処理結果を高知識オペレータ端末1803に送り、上記以外の文書を一般オペレータ端末1804に送る。

#### 【0069】

一般にはあまり発生しないケースであるが、回答例文候補検定1701で全ての重要部の1位の回答例文候補が確かに正しいと判定され、且つ回答新規作成判定1801で一つでも回答新規作成する必要があると判定された重要部が存在する文書は高知識オペレータ端末に送る。一般オペレータ端末では、図15に示すような回答作成画面を用いて、オペレータによるキーボードやマウス操作により、重要部毎に回答例文候補から適切な回答例文が選択されることにより、回答が作成される。そしてオペレータの操作により、作成された回答は決裁者端末に送られる。ただし回答例文候補では対応できない重要部が存在した文書は、高知識オペレータ端末に送られる。振り分け先を決定するにあたっては、上記の回答例文候補信頼度や回答新規作成度以外の評価値で、作成された回答の確からしさを表すものを用いてもよい。

#### 【0070】

高知識オペレータ端末では、図15に示すような回答作成画面を用いて、オペレータによるキーボードやマウス操作により、重要部毎に回答例文で対応できる部分は回答例文候補の中から適切な回答例文が選択され、回答例文で対応できない部分は新規に回答が作成される。そして作成された回答はオペレータの操作により決裁者端末に送られる。

#### 【0071】

決裁者端末では、図15に示すような回答作成画面を用いて、決裁者により、重要部毎に回答が正しいかどうかを確認され、すべて正しいと確認されると、質問者に回答が送られる。また回答に誤りがある場合には、決裁者により訂正される、或いは高知識オペレータ端末、或いは一般オペレータ端末に再度質問文書を送る。

#### 【0072】

本実施例では、回答候補検定と回答新規作成の両方を行い、決済者端末、高知識オペレータ端末、一般オペレータ端末に質問文書と自動振り分け処理結果を振り分ける場合を説明したが、次の2つの構成にしても良い。

#### 【0073】

図17に示すように、回答候補検定のみを行い、振り分けA1702にて、全ての重要部の1位の回答候補が確かに正しいと判定した文書と一連の処理結果を決裁者端末に送り、それ以外をオペレータ端末1704に送るという構成にしても良い。

#### 【0074】

図18に示すように、回答新規作成判定のみを行い、振り分けB1802にて、一つでも回答新規作成する必要があると判定された重要部が存在する文書と一連の処理結果を高知識オペレータ端末1803に送り、それ以外を一般オペレータ端末1804に送ると言う構成にしても良い。

#### 【0075】

自動振り分け処理を行うに伴い追加された回答例文候補検定1701、回答新規作成検定1801、振り分けA1702、振り分けB1802、振り分けC1901、振り分けD2005の処理を行うプログラムは外部装置1304やメモリ1305等の記憶装置に格納され、CPUによって実行される。また回答例文候補検定、回答新規作成判定に用いる閾値の銃砲は、外部装置1304やメモリ1305の記憶装置に格納され、図18、図19、図20に示す一連の処理に従って、CPUによりロード/ストアされる。

#### 【0076】

これにより、従来すべての文書に対して重要部毎に、回答選択作業、或いは回答新規作

成を行わなければならなかったが、直接決済者端末に送られる文書は、回答候補の確認作業が行われるのみとなり、業務が効率化される。またオペレータの数が削減されるためコストが削減される。そして、一般のオペレータにより、回答を選択するだけで済む質問文書の回答作成作業が行われ、高知識オペレータにより、回答新規作成を必要とする質問文書の回答作成作業が行われるため、業務が効率化される。そして、一般的に賃金の高い高知識オペレータの数を削減でき、コスト削減できる。

#### 【0077】

以上が自動振り分け機能を追加した場合の処理の説明である。本システム構成は、決済者端末における確認作業をなくし、決裁者に送られる回答がそのまま質問者を送られるシステム構成にもできる。

#### 【0078】

上記自動振り分け処理において、重要区間の抽出処理の構成を変更し、図20の構成にすることができる。

#### 【0079】

これは上記構成中の重要区間の抽出処理において、誤りはあるが全ての重要区間を抽出する重要区間抽出R12001を行い、全ての重要区間を抽出できないが、全て正しい重要区間を抽出する重要区間抽出P12002を行い、重要部抽出R1の結果と重要部抽出P1の結果が等しいかどうかを判定し2003、重要区間抽出R1の結果と重要区間抽出P1の結果が等しくない場合のみ、重要区間抽出R1の結果と重要部抽出P1の結果を消去し、重要区間抽出RN1604を行う。

#### 【0080】

このとき誤りはあるが全ての重要区間を抽出する重要区間抽出R1は、実施例1で説明した重要区間抽出方法（図2～図4）を用いる。ただし、重要区間の抽出の際に用いる閾値をすべての重要区間が抽出できるような低い値に設定しておく。

#### 【0081】

全ての重要区間を抽出できないが、全て正しい重要区間を抽出する重要区間抽出P1は、実施例1で説明した重要区間抽出方法（図2～図4）を用いる。ただし、重要区間の抽出の際に用いる閾値をすべての非重要部が抽出されない高い値に設定しておく。

#### 【0082】

重要区間抽出RNは、実施例1で説明した重要区間抽出方法（図2～図4）を用いる。ただし、重要区間の抽出の際に用いる閾値を非重要部の抽出率が少なく、多くの重要部が抽出されるような値に設定しておく。

#### 【0083】

次に、重要区間の結合104を行う。次に、重要部毎の回答例文候補生成105を行う。次に、回答例文候補検定1701を行う。次に、回答新規作成判定1801を行う。そして、振り分けD2005を行う。

#### 【0084】

ここでは、重要区間抽出R1の結果と重要区間抽出P1の結果が等しい場合のみ、回答例文候補検定で全ての重要部の1位の回答例文候補が確かに正しいと判定した文書を決裁者端末に送り、回答新規作成判定で一つでも回答新規作成する必要があると判定された重要部が存在する文書を高知識オペレータ端末に送る。そして、それ以外の場合は、一般オペレータに文書を送る。

#### 【0085】

この構成にすることにより、決済者端末や高知識オペレータ端末に振り分けられる文書の数が減ってしまうが、質問文書中の重要区間抽出の結果に漏れが少くなり、さらにつれての重要区間の抽出結果に誤りが少なくなるため、誤りの少ない振り分けを行うことができる。

#### 【0086】

本出願に係る代表的な発明として、上述のほかにも、例えば以下のようなものがある。

#### 【0087】

質問文書の入力を受け付ける入力装置と、複数の回答例文および上記回答例文のそれについて該回答例文に対応する質問文書に含まれるキーワードの頻度を示す質問文書キーワード頻度情報を記憶する記憶装置と、上記質問文書の重要区間に対応する回答例文を該複数の回答例文から選択するCPUとを有する質問応答装置に、上記入力された質問文書を複数の区間に区切るステップと、上記質問文書キーワード頻度情報を用いて、上記複数の区間の各区間にについて、該区間が上記記憶されている複数の回答例文のそれに対応する質問内容である尤度を求めるステップと、上記各区間にについて求められた尤度のいずれかと予め定められた閾値との関係に基づいて上記重要区間を抽出するステップとを実行させるためのプログラム。

#### 【0088】

質問文書の入力を受け付ける入力装置と、複数の回答例文および質問文書の重要部に用いられるキーワードについての重要部キーワード頻度情報または質問文書の非重要部に用いられるキーワードについての非重要部キーワード頻度情報を記憶する記憶装置と、上記質問文書の重要区間に対応する回答例文を該複数の回答例文から選択するCPUとを有する質問応答装置に、上記入力された質問文書を複数の区間に区切るステップと、上記重要部キーワード頻度情報または上記非重要部キーワード頻度情報を用いて、上記複数の区間の各区間の重要度を計算するステップと、上記計算された重要度が所定の閾値を超える区間を重要区間として抽出するステップとを実行させるためのプログラム。

#### 【0089】

質問文書と、該質問文書から抽出された複数の重要部に関する重要部情報と、該複数の重要部のそれに対応する回答例文候補に関する回答例文候補情報の入力を受ける入力部と、情報処理部と、該質問文書に対する回答文書を表示する表示部と、該表示された回答例文候補に関するユーザ入力を受け付けるユーザ入力装置を有し、上記表示部は、上記重要部情報で特定される複数の重要部の箇所を第1の強調表示で強調表示した上記質問文書と、上記回答例文候補情報に示される回答例文候補を含む回答文書とを表示することを特徴とする回答作成装置。

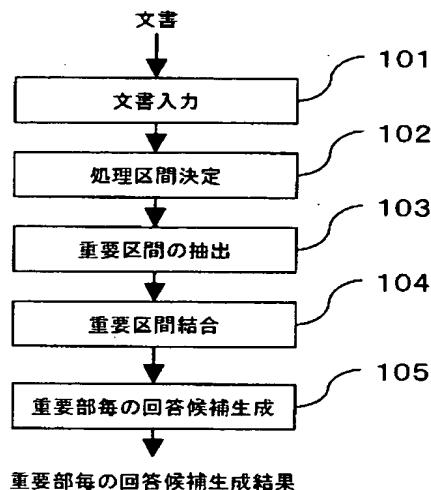
#### 【図面の簡単な説明】

#### 【0090】

- 【図1】重要部抽出による複数相談対応処理を表す図。
- 【図2】重要区間抽出処理を表す図（回答例文候補尤度利用の場合）。
- 【図3】重要区間抽出処理を表す図（区間重要度利用の場合）。
- 【図4】重要区間抽出処理を表す図（回答例文候補尤度と区間重要度利用の場合）。
- 【図5】入力文書例を示す図。
- 【図6】読点/改行を区切りとした区間設定を示す図。
- 【図7】重要区間抽出例を示す図。
- 【図8】重要区間毎の回答候補生成を示す図。
- 【図9】重要区間結合結果を示す図。
- 【図10】重要部抽出による回答候補生成結果を示す図。
- 【図11】回答候補生成処理に用いる回答例文毎のキーワード頻度情報を示す図。
- 【図12】重要部の判定に用いる重要部/非重要部キーワード頻度情報を示す図。
- 【図13】装置を表す図。
- 【図14】自動振り分け処理システムを表す図。
- 【図15】回答作成画面を示す図。
- 【図16】重要部のみを表示した回答作成画面を示す図。
- 【図17】回答例文候補検定を加えた自動振り分け処理を表す図。
- 【図18】回答新規作成判定を加えた自動振り分け処理を表す図。
- 【図19】回答例文候補検定と回答新規作成判定を加えた自動振り分け処理を表す図
- 。 【図20】3種類の重要部抽出を用いた自動振り分け処理を表す図。

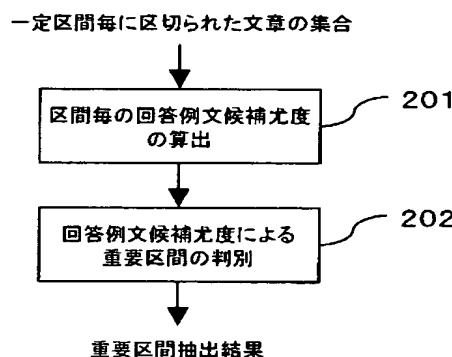
【書類名】図面  
【図1】

図1



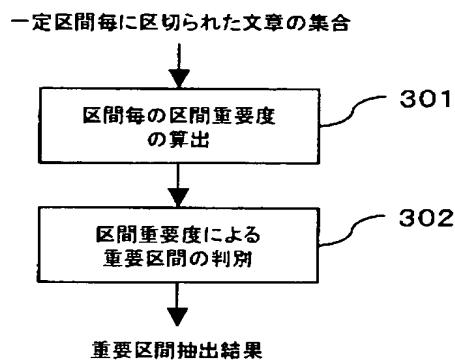
【図2】

図2



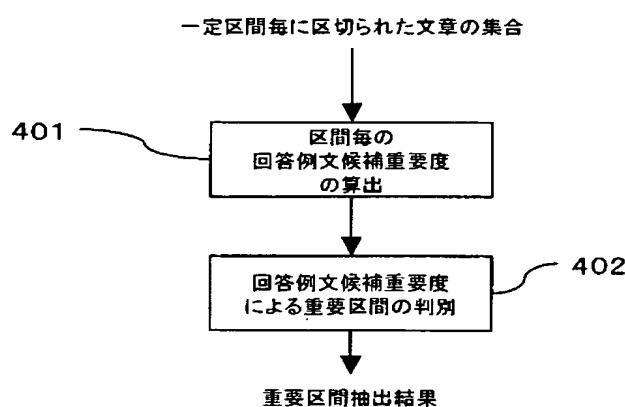
【図3】

図3



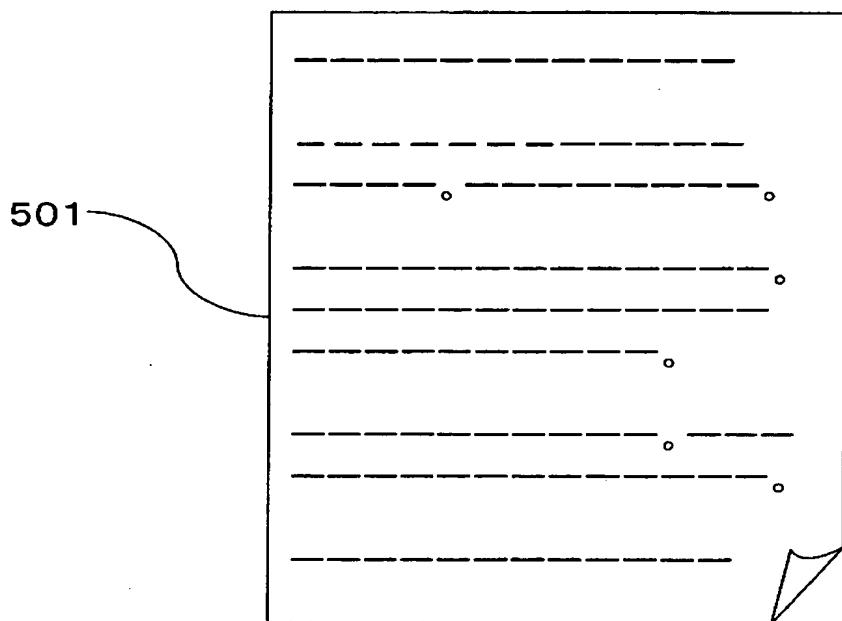
【図4】

図4



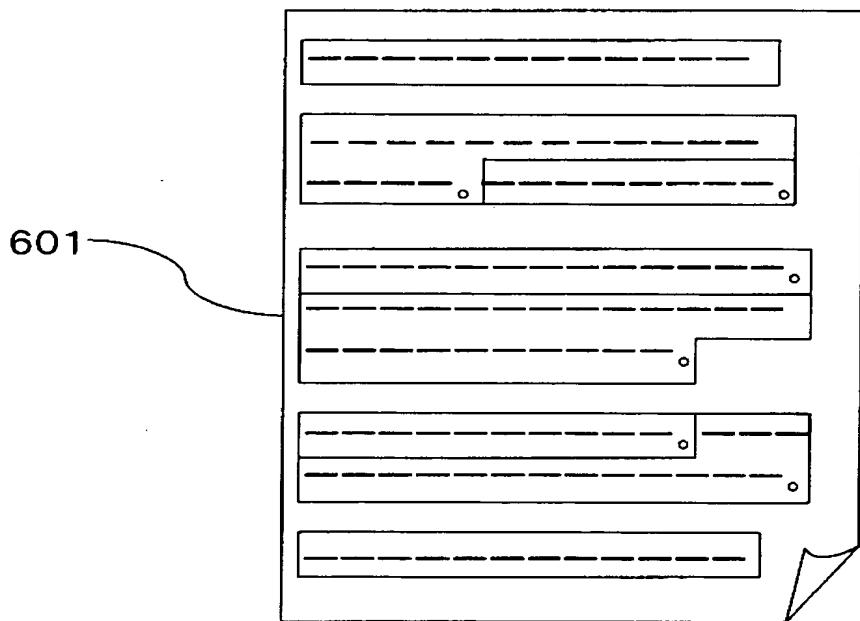
【図5】

図5



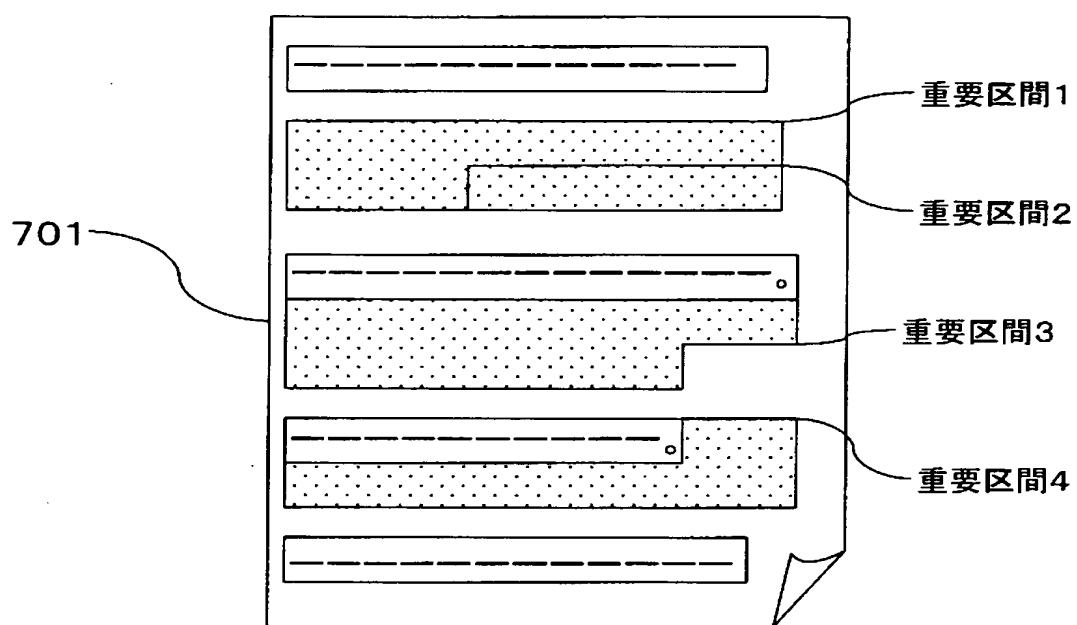
【図6】

図6

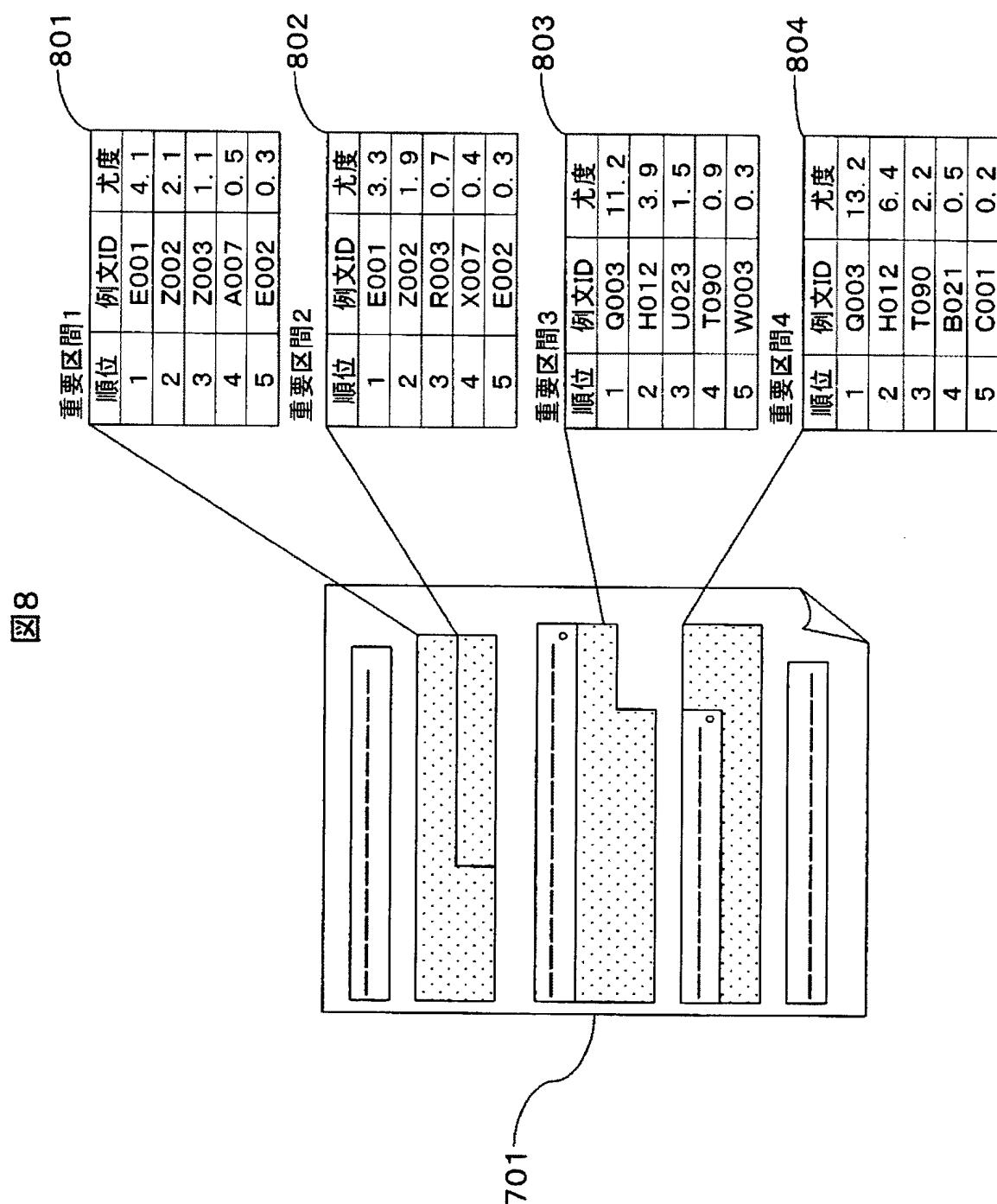


【図7】

図7

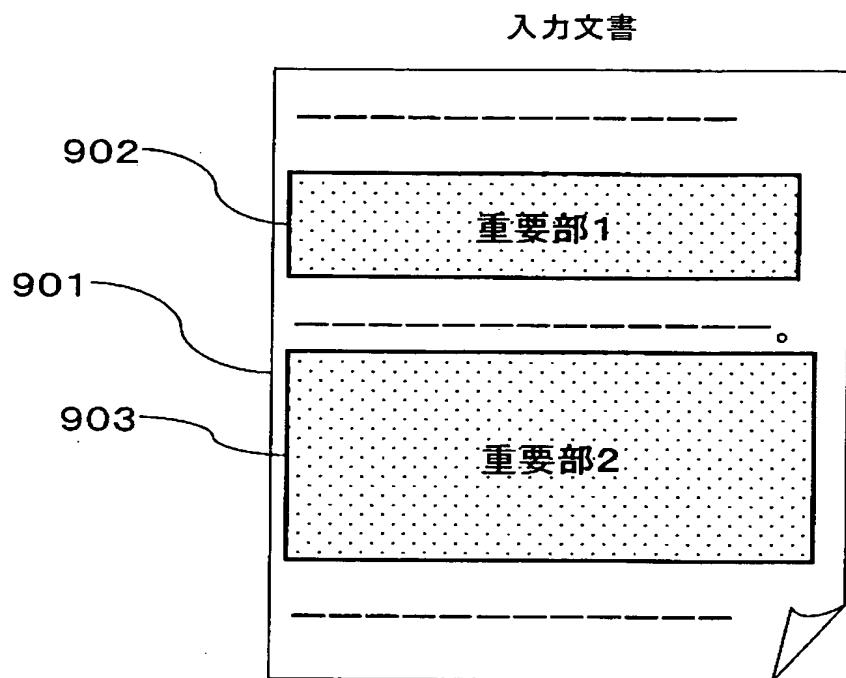


【図8】



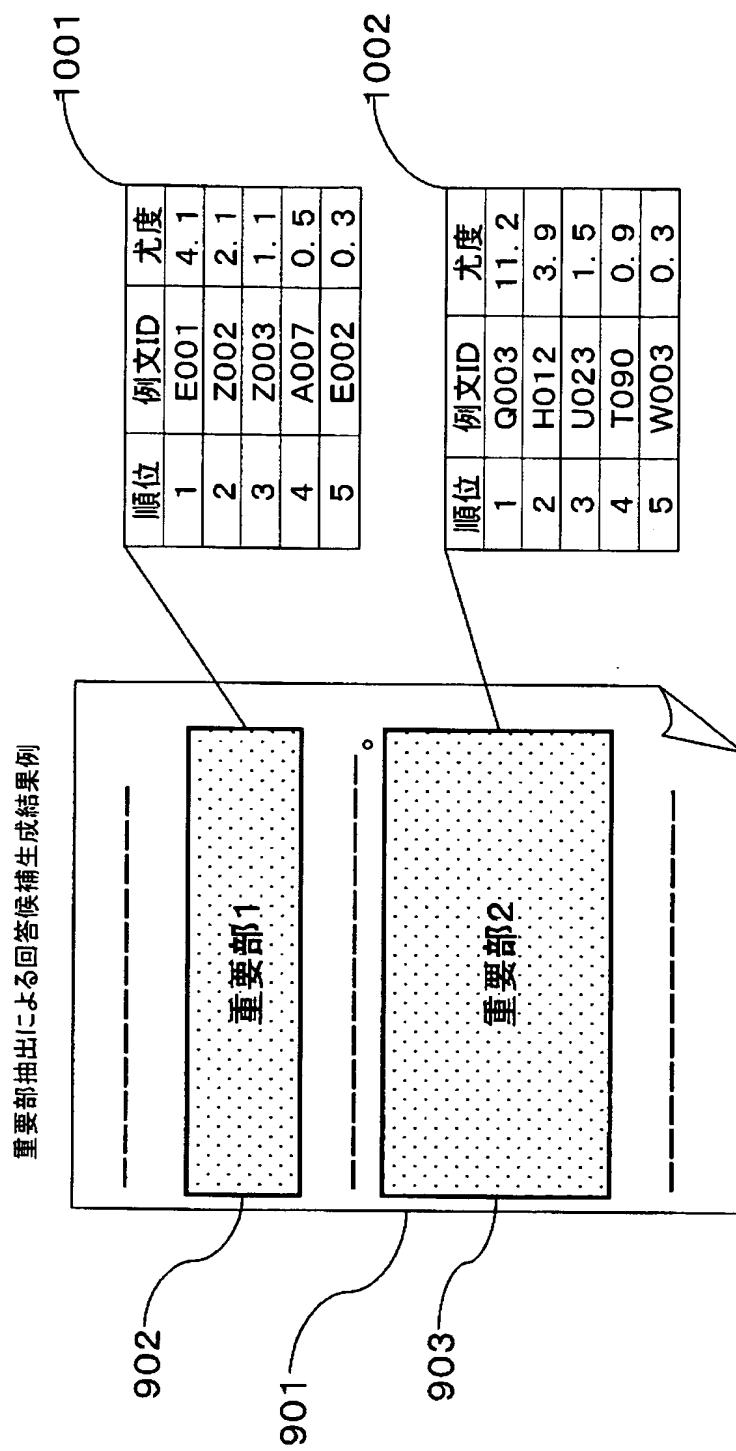
【図9】

図9

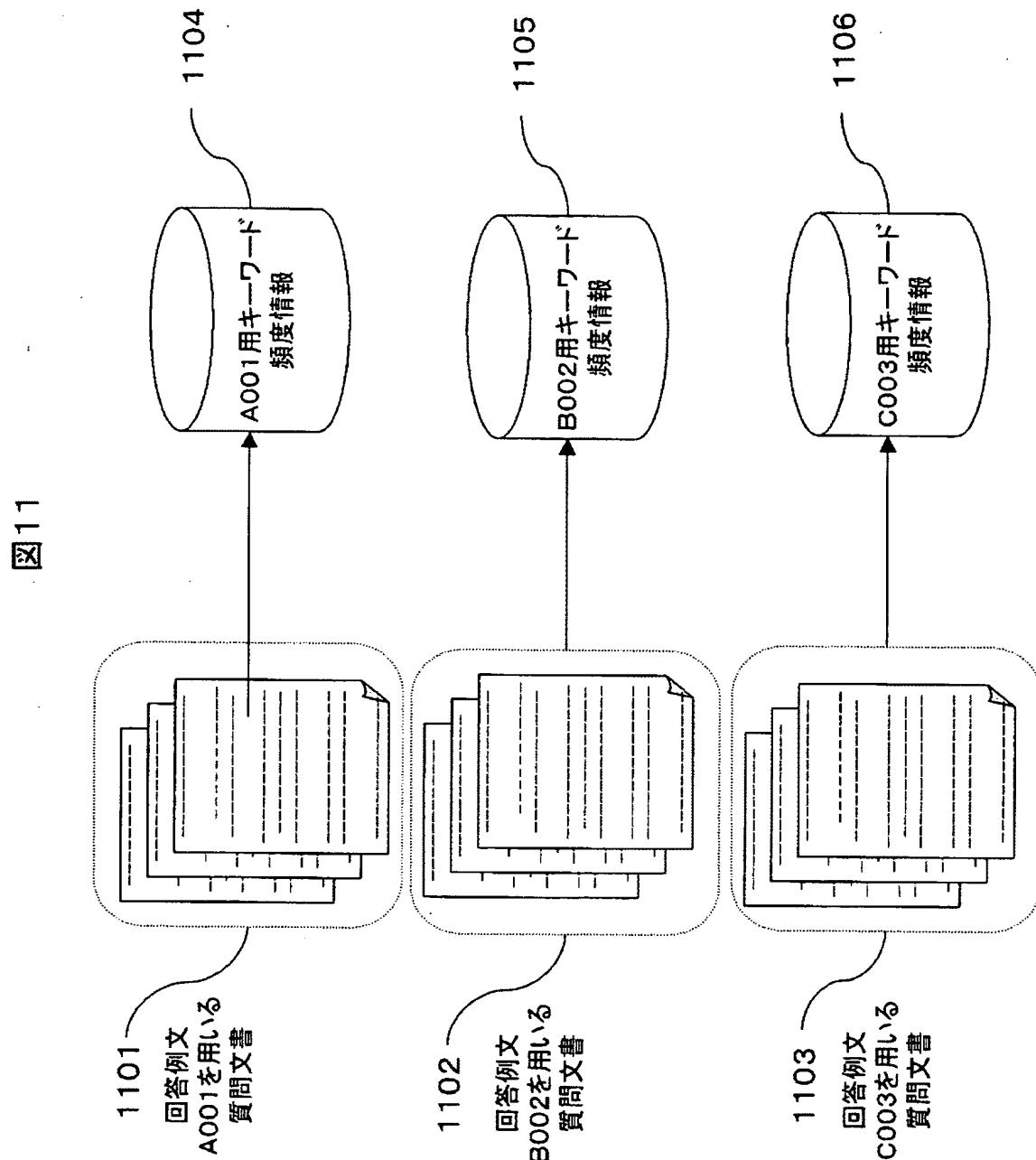


【図10】

図10

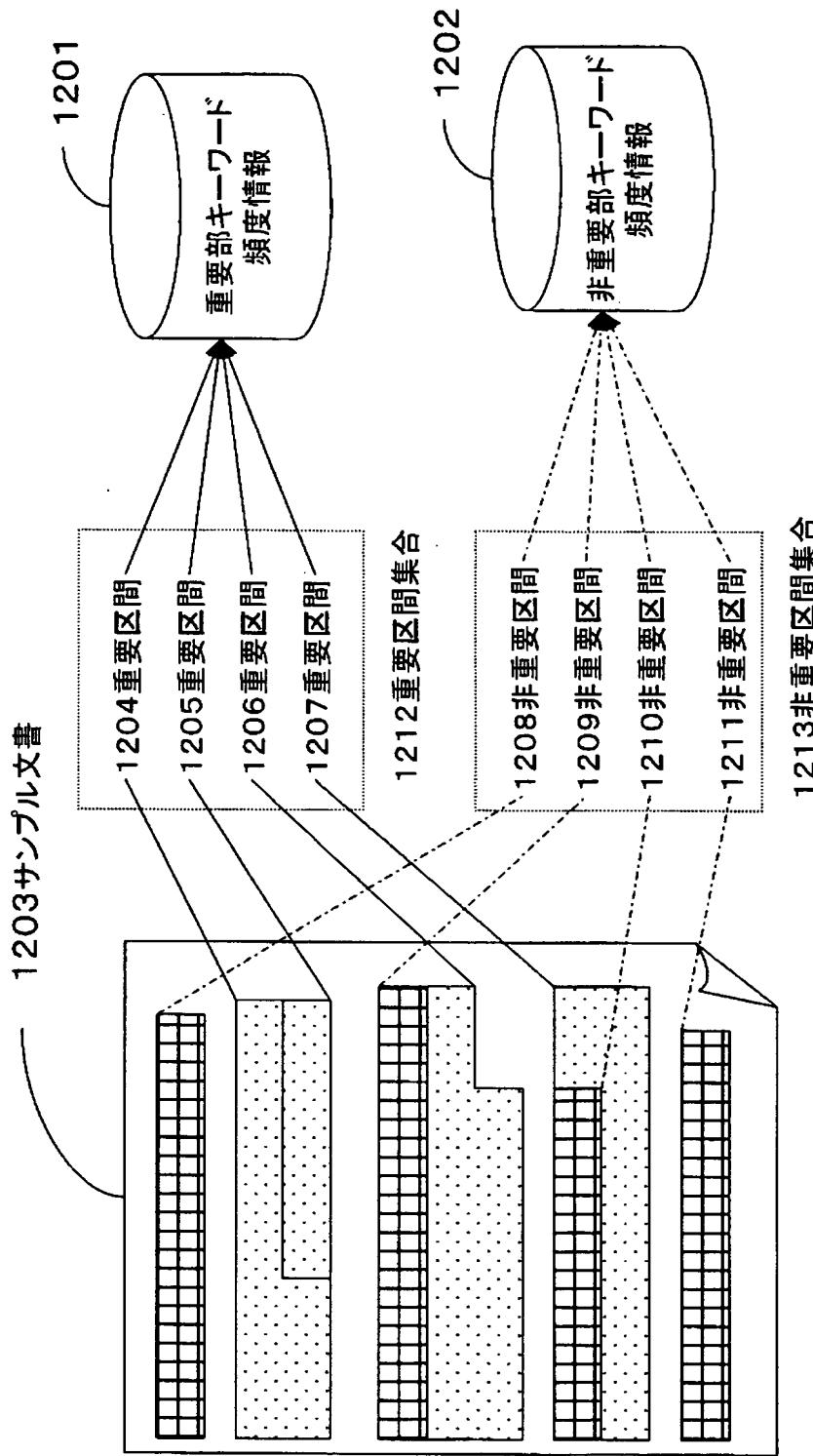


【図11】



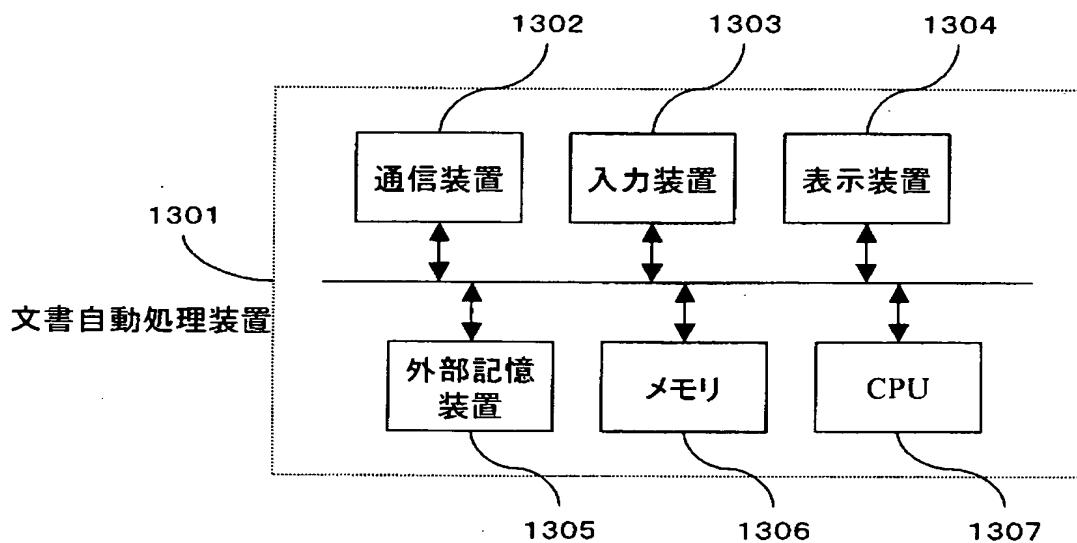
【図12】

図12



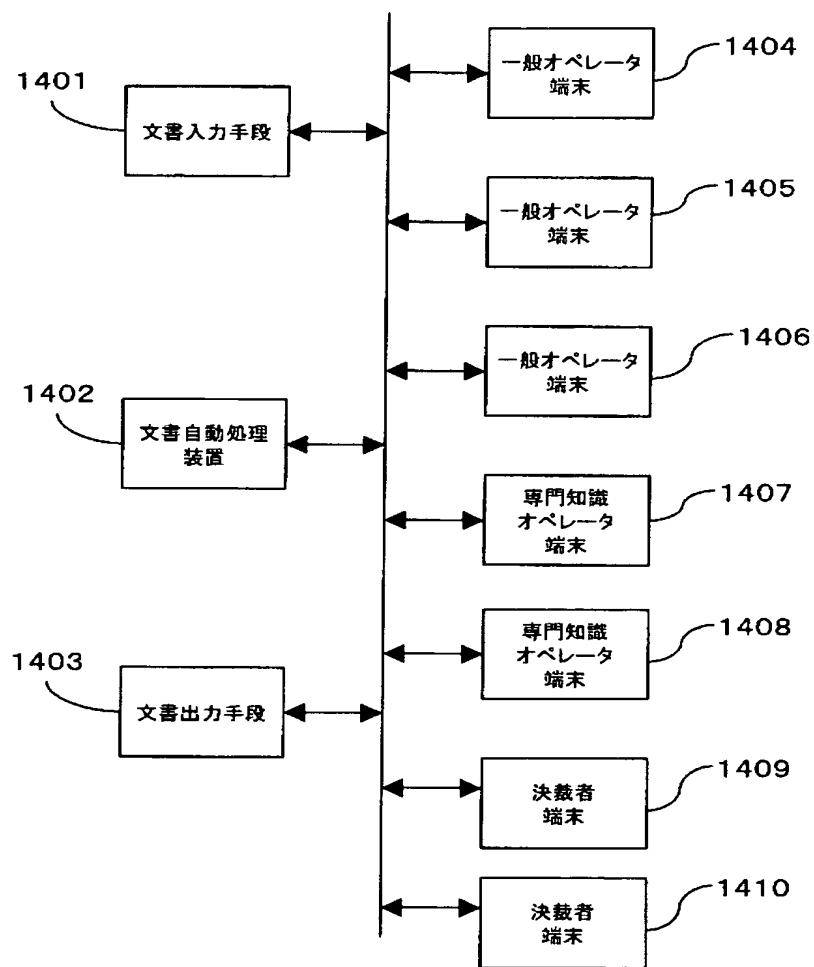
【図13】

図13



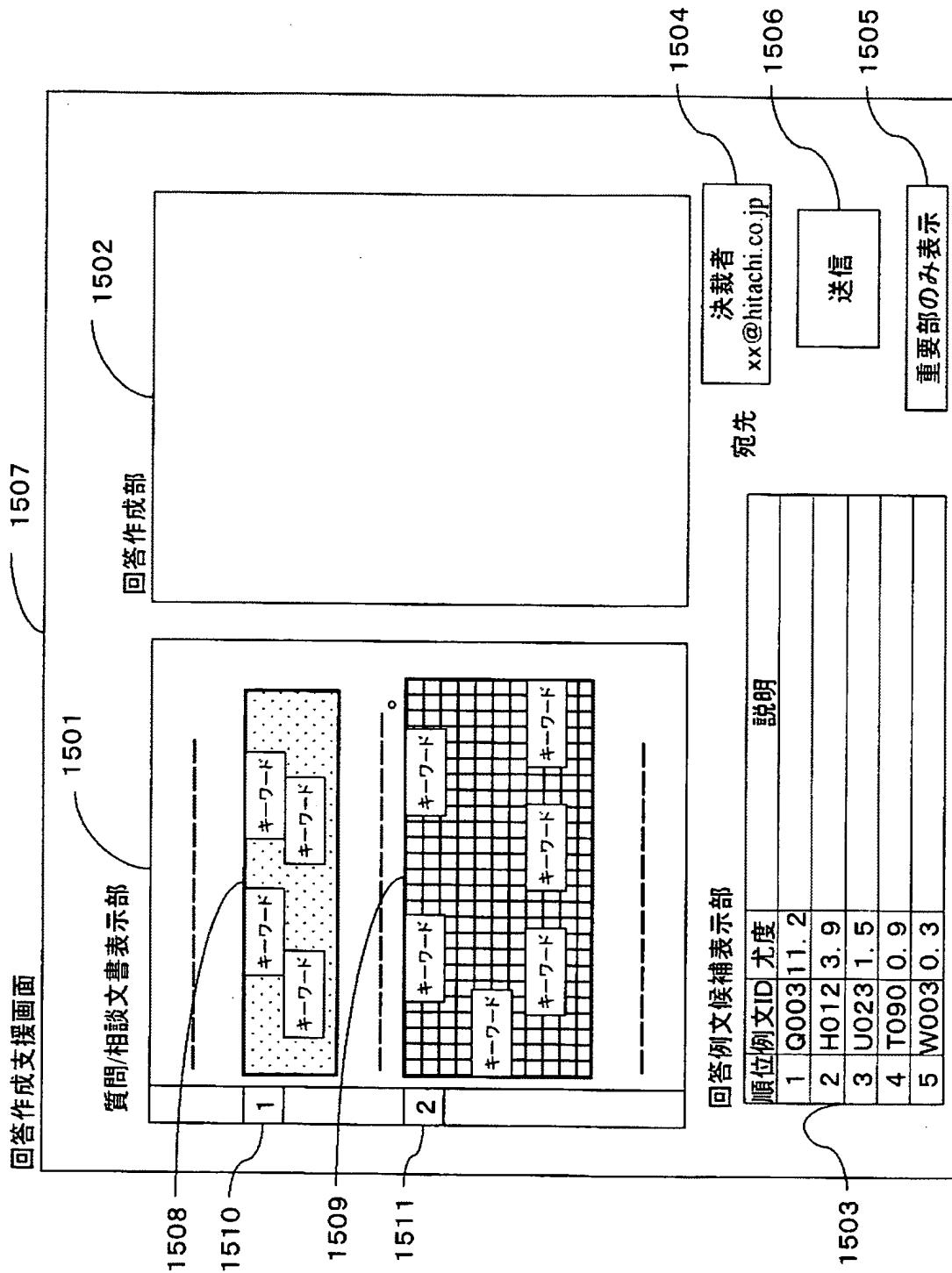
【図14】

図14

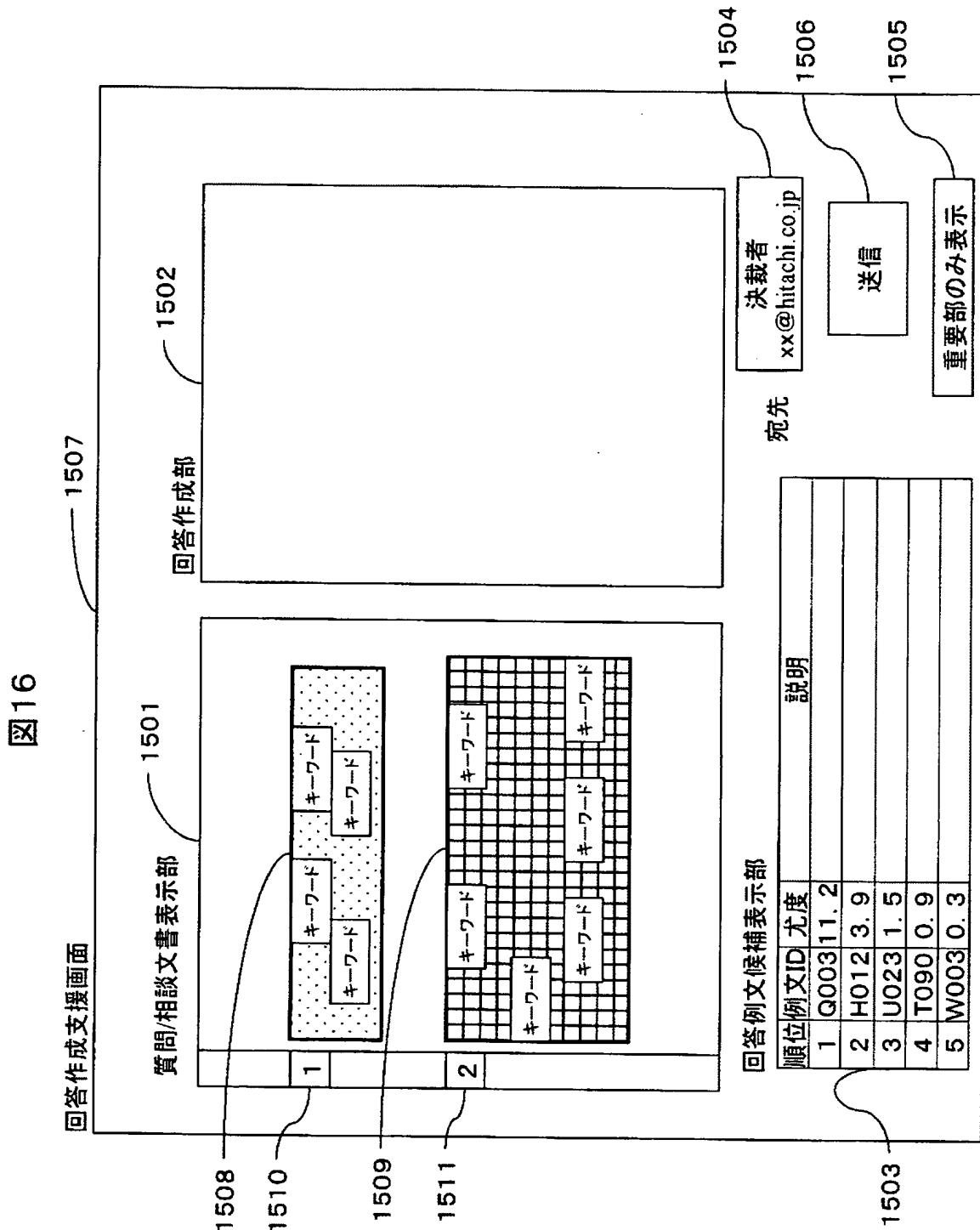


【図15】

図15

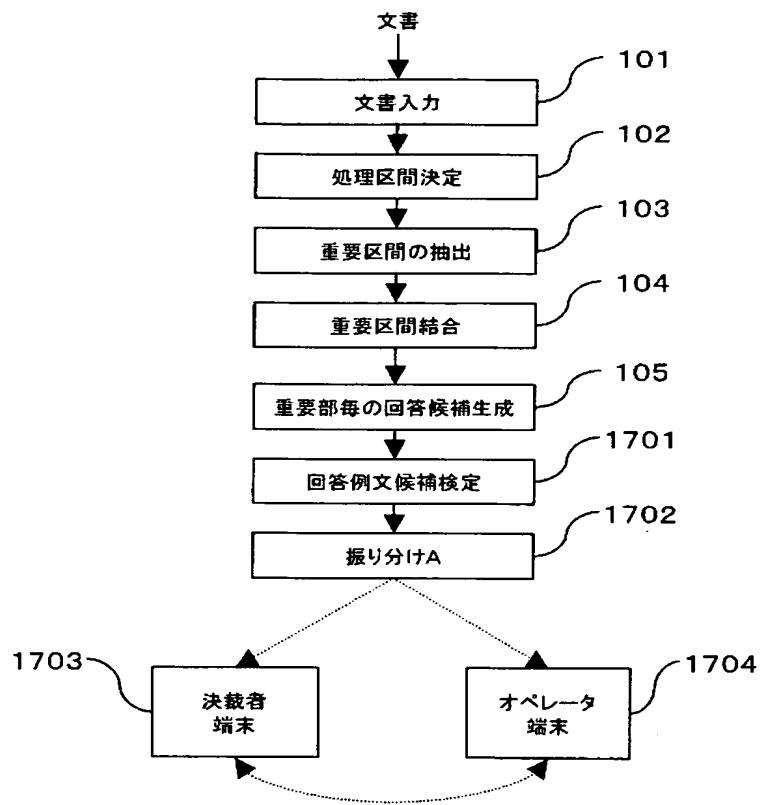


【図 16】



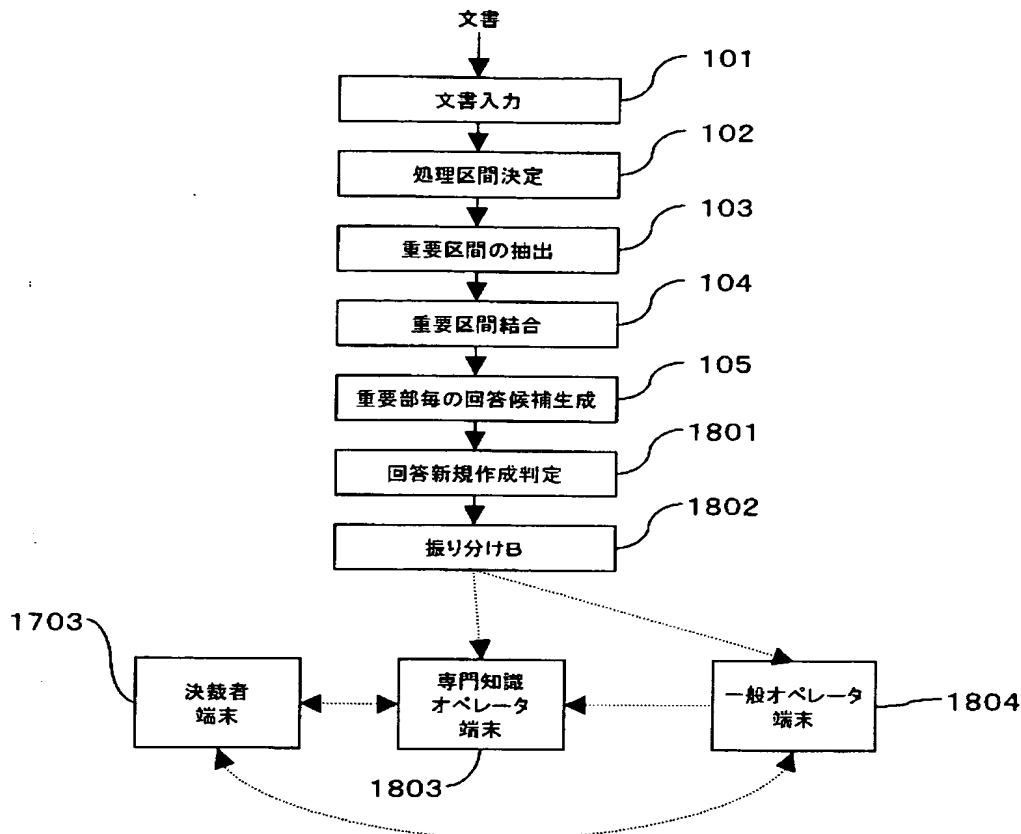
【図17】

図17



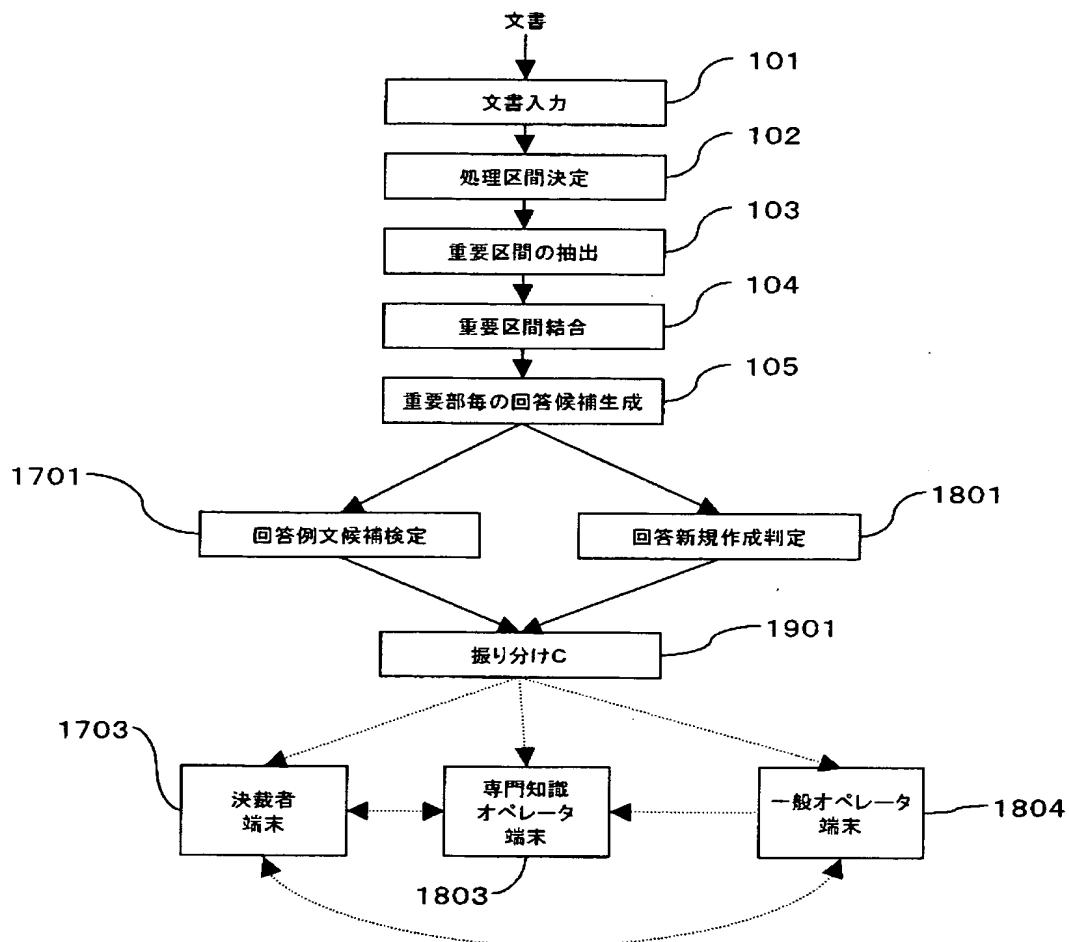
【図18】

図18



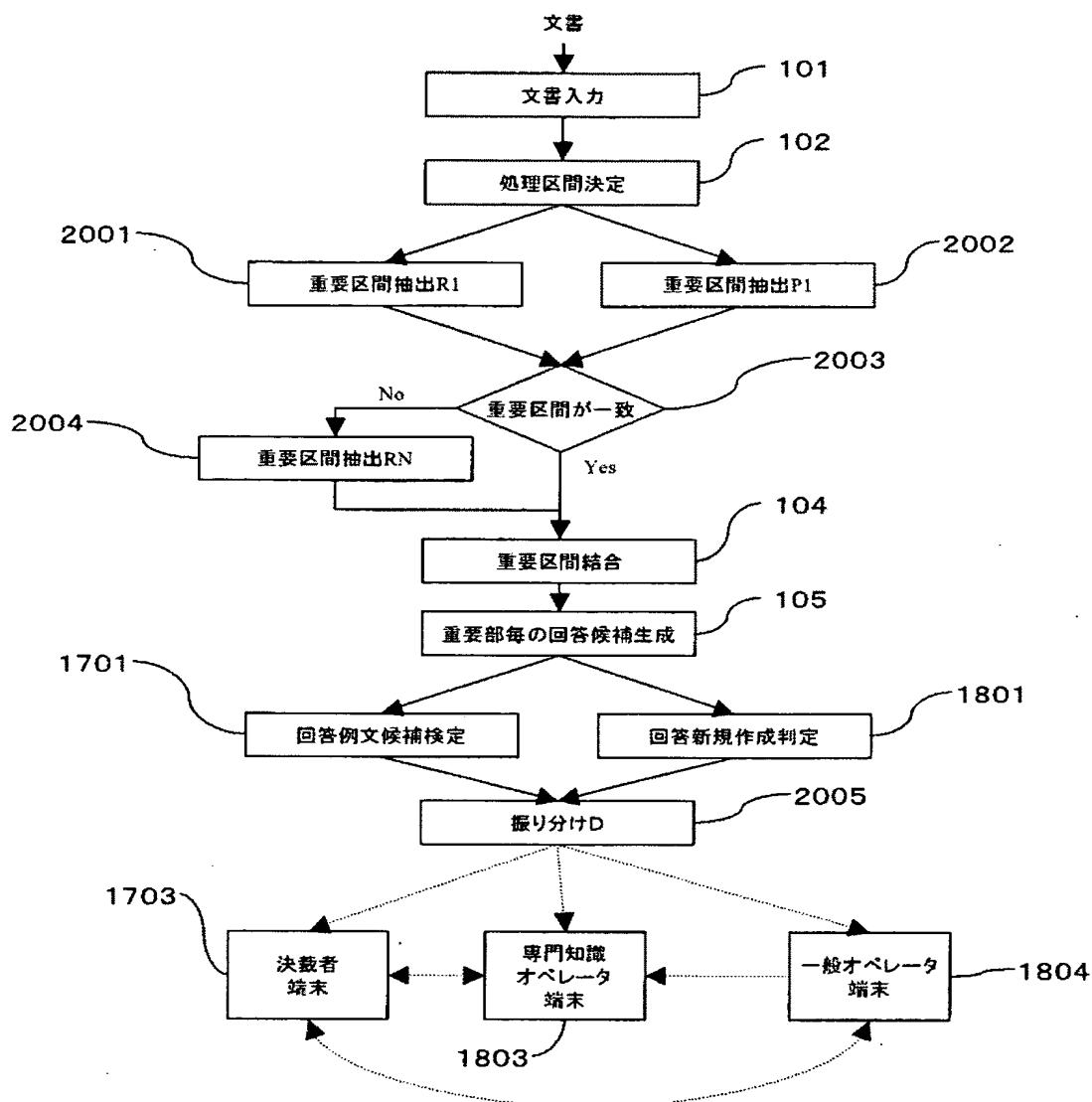
【図19】

図19



【図20】

図20



【書類名】要約書

【要約】

【課題】 1つの文書内に複数の質問内容がある場合にも、それぞれの質問に対して精度良く回答例文候補を生成し、効率的な回答作成作業を行うことが可能な質問応答方法および質問応答装置を提供することを目的とする。

【解決手段】 質問文書を所定の区間に区切り、区切られた区間毎に重要であるか否かを判定することで重要区間を抽出し、重要区間毎に各回答例文に対する質問内容であるかの度合を表す回答例文候補尤度を算出し、回答例文候補尤度を用いて意味的に近い重要区間を結合することで質問内容毎に重要部を抽出し、重要部毎に予め用意された回答例文から回答例文候補を選択する。また、回答例文候補の確からしさの度合を示す回答例文候補信頼度と、回答新規作成が必要か否かの度合を示す回答新規作成度を算出し、これらの値を用いて異なるオペレータ端末に質問文書を振り分ける。

【選択図】 図15

**認定・付加情報**

特許出願の番号	特願2003-320602
受付番号	50301512981
書類名	特許願
担当官	第七担当上席 0096
作成日	平成15年 9月16日

## &lt;認定情報・付加情報&gt;

【提出日】 平成15年 9月12日

特願 2003-320602

出願人履歴情報

識別番号 [000005108]

1. 変更年月日 1990年 8月31日

[変更理由] 新規登録

住所 東京都千代田区神田駿河台4丁目6番地  
氏名 株式会社日立製作所